Research Article

Compositional Data Analysis for Modelling and Forecasting Mortality with the α-Transformation

Han Ying Lim[1], **Dharini Pathmanathan**[1,2,3], **Sophie Dabo-Niang**[4,5]

1. Institute of Mathematical Sciences, Faculty of Science, Universiti Malaya, Malaysia; 2. Universiti Malaya Centre for Data Analytics, Universiti Malaya, Malaysia; 3. Center of Research for Statistical Modelling and Methodology, Universiti Malaya, Malaysia; 4. UMR8524-Laboratoire Paul Painlevé, Inria-MODAL, University of Lille, CNRS, France; 5. CRM-CNRS, Université de Montréal, Canada

Mortality forecasting is crucial for demographic planning and actuarial studies, especially for projecting population ageing and longevity risk. Classical approaches largely rely on extrapolative methods, such as the Lee-Carter (LC) model, which use mortality rates as the mortality measure. In recent years, compositional data analysis (CoDA), which respects summability and non-negativity constraints, has gained increasing attention for mortality forecasting. While the centred log-ratio (CLR) transformation is commonly used to map compositional data to real space, the α-transformation, a generalisation of log-ratio transformations, offers greater flexibility and adaptability. This study contributes to mortality forecasting by introducing the α-transformation as an alternative to the CLR transformation within a non-functional CoDA model that has not been previously investigated in existing literature. To fairly compare the impact of transformation choices on forecast accuracy, zero values in the data are imputed, although the α-transformation can inherently handle them. Using age-specific life table death counts for males and females in 31 selected European countries/regions from 1983 to 2018, the proposed method demonstrates comparable performance to the CLR transformation in most cases, with improved forecast accuracy in some instances. These findings highlight the potential of the α-transformation for enhancing mortality forecasting within the non-functional CoDA framework.

Corresponding author: Dharini Pathmanathan, dharini@um.edu.my

 doi.org/10.32388/Z69TVG.2

1. Introduction

Mortality forecasting plays a crucial role in demographic analysis and informs strategic planning across sectors such as healthcare, insurance and social welfare. Accurate projections of population ageing and longevity risk are essential for designing sustainable pension systems and effectively pricing longevity-linked financial products [1][2]. Recognising the importance of anticipating human mortality and longevity, the development of mortality forecasting methods dates back to the twentieth century [3].

In general, mortality forecasting methods are categorised into expert judgement, extrapolative methods and epidemiological models [4]. Among these, the Lee-Carter (LC) model which forecasts mortality based on log mortality rates has gained popularity since its establishment [5]. It has been extended into different variants to improve forecast accuracy over the decades [6].

Widely regarded as a benchmark in mortality forecasting, the LC model employs a statistical time series approach to project a single time-varying parameter for forecasting mortality rates. This process is purely extrapolative with minimal subjective judgment [6]. The time index κ_t and age pattern β_x are estimated using singular value decomposition (SVD) applied to a centred matrix of log mortality rates, enabling the projection of life table death densities [7][5][8]. This approach leverages the approximately log-linear decline in age-specific mortality rates over time and allows the use of multivariate statistical techniques designed for unbounded variables [9].

In mortality forecasting, various mortality measures can be used, each influencing the choice of modelling methods and the resulting forecasts. While most studies commonly model mortality rates [7][5][10][8], others have focused on alternative measures such as death probabilities [11][1][12], life table deaths [13][14][9][15], survival probabilities [16] and life expectancy at birth [17]. Bergeron-Boucher et al. [18] observed that while death rates and death probabilities generally yield similar trends, both measures often lead to more pessimistic forecasts compared to survival probabilities, life table deaths and life expectancy. However, this does not imply that death rates or death probabilities consistently offer higher accuracy, particularly when forecasts are data-dependent [18]. Thus, the choice of measure should ultimately depend on the research context, including the study objective, research question and target population [18]. For instance, when the focus is on the age distributions of mortality, life expectancy may not be an appropriate choice. In recent years, age-at-death distributions have gained increasing attention for their ability in capturing mortality conditions [15], central measures of longevity [19] and lifespan variability [20].

Compositional data analysis (CoDA) is an analytical framework designed to handle compositional data that are positive vectors carrying relative information, such as proportions, that represent parts of a whole with a fixed sum [21]. Such data is commonly found in geochemistry and atmospheric science. Transformations are needed to map the compositional data from the Aitchison simplex to the real space before conducting standard statistical analyses [21]. The centred log-ratio (CLR) transformation is widely used due to its interpretability and ability to preserve distances [21].

Since life table death counts $d_{t,x}$ are non-negative, range between 0 and the life table radix, and naturally sum to the radix each year [9], they can be treated as compositional data. Forecasting $d_{t,x}$ using a log-linear approach often results in predicted values that vary independently across ages and fail to preserve the life table radix constraint [18]. This limitation can be addressed by leveraging the constant sum constraint inherent in the CoDA framework, which induces a natural covariance structure among components [18].

Oeppen's [9] pioneering work introduced a CoDA-based framework to mortality forecasting, analogous to the LC model, that focuses on forecasting the redistribution of the density of $d_{t,x}$. Within this framework, deaths are progressively redistributed from younger to older ages. Oeppen [9] found that the multiple-decrement compositional forecasts by age and cause are not necessarily more pessimistic than single-decrement forecasts by age alone, thereby contradicting earlier findings based on mortality rates [22].

Bergeron-Boucher et al. [13] then extended the CoDA model for regional coherent mortality forecasting, akin to the Li-Lee model [8]. Their findings highlight that both coherent and non-coherent CoDA models yield less biased forecasts with increased accuracy for many selected countries compared to their LC-based counterparts. This improvement is partly attributed to the use of $d_{t,x}$ as mortality measure and the application of the CLR transformation, which accounts for the changing rate of mortality improvement over time [13]. Furthermore, the summability constraint in compositional data also preserves coherence across populations, thereby addressing one of the key limitations of the LC model [13].

Acknowledging the strength of CoDA framework in capturing dependencies between causes of death, Kjærgaard et al. [14] proposed two CoDA-based models to forecast cause-specific death distributions within a single population. Subsequently, Shang and Haberman extended the CoDA framework to a functional setting by introducing a functional CoDA model [23] that adapts the Hyndman-Ullah (HU) model [7], followed by a weighted functional CoDA model [15]. These studies have shown that CoDA-based approaches can improve the accuracy of mortality forecasting.

However, these studies use log-ratio approaches to transform $d_{t,x}$, which have major drawbacks as they lack flexibility and cannot handle zeros due to their logarithmic nature. These limitations can be addressed by the α-transformation introduced by Tsagris et al. [24]. The α-transformation generalises log-ratio transformations, offering greater flexibility through the α-parameter [24]. With the α-parameter commonly ranges between 0 and 1 [25][26], it balances Euclidean data analysis (EDA) and log-ratio analysis (LRA). The intermediate α values sometimes outperform $\alpha = 0$ (LRA) and $\alpha = 1$ (EDA) [27]. This flexibility allows the α-transformation to be applied to data containing zeros using strictly positive values of α [24][27][28]. Although the α-transformation does not satisfy all the theoretical properties outlined by Aitchison [21], such as scale invariance, perturbation invariance and subcompositional dominance, its practical applicability remains unaffected since these properties primarily support the LRA methods [24][28].

Numerous empirical studies have demonstrated that the α-transformation can enhance performance in both regression [29] and classification tasks [27]. In the context of forecasting, Shang and Haberman [26] found that the α-transformation outperformed log-ratio transformations within the functional CoDA framework in short-term forecasting for Australian mortality data. These findings are consistent with those of Giacomello [25], who extended the CoDA framework by applying a multivariate functional α-transformation to provincial mortality data in Italy.

The research gap lies in the unexamined application of the α-transformation within the CoDA framework under a non-functional data setting for all-cause mortality forecasting, by treating age as discrete rather than a continuum. This paper addresses the gap by evaluating forecast performance across multiple countries, highlighting the potential of the α-transformation to produce better or at least comparable results relative to the CLR transformation.

Section 2 provides a detailed description of methodology, covering the key steps and analytical framework employed in this study. Subsequently, Section 3 presents the results, along with a comprehensive discussion revolving around the forecast accuracy of each transformation. Lastly, Section 4 concludes the paper by summarising key findings and suggesting possible extensions for future research.

2. Methodology

This study comprises several phases, including data preprocessing, α-parameter tuning, modelling and forecasting, and model evaluation. All analyses are performed using R Statistical Software version

4.1.1 [30].

2.1. Data preprocessing

Observed mortality data for males and females from age 0 to an open interval 110+ from 31 selected European countries/regions, covering the period 1983 to 2018, are retrieved from the Human Mortality Database (HMD) [31] using the `demography` package in R [32]. The countries/regions are selected to maximise data completeness while ensuring a common timeframe. The pre-pandemic period is chosen to avoid anomalies and uncertainties introduced by COVID-19, as forecast performance is known to be highly sensitive to data quality and stability. Zero values are imputed to ensure a fair comparison between the α and CLR transformations.

Similar to the pipeline of Bergeron-Boucher et al. [13], observed death counts D_x for each country/region are first calculated based on observed mortality rates M_x and exposure-to-risk estimates E_x. However, at older ages above 80, M_x values often exhibit considerable random variation due to unboundedly high rates, smaller denominators in E_x or measurement error [33][6]. To address this, the Kannisto model [34] is applied to smooth mortality rates for ages 80 and above, separately for males and females [33]. The model uses a Poisson log-likelihood procedure, where a logistic curve better fits old-age mortality patterns compared to alternative models [33]. As a result, zeros and missing values are eliminated at advanced ages.

For ages below 80, zeros are present for some specific years in some countries/regions, which can lead to undefined results in the CLR transformation due to the nature of logarithm. A multiplicative replacement strategy [35] is therefore applied to D_x to impute zeros [13]. This non-parametric method is coherent with simplex operations and retains the covariance structure of non-zero components, ensuring minimal distortion to the overall mortality pattern [35], while enabling the application of log-ratio transformations by replacing zeros with small positive values. In this study, although the α-transformation can handle zeros, zero replacement is still necessary to ensure a fair comparison of its impact on forecast accuracy against the CLR transformation.

Basically, a composition $\mathbf{x} = [x_1, x_2, \dots, x_D]$ of D_x containing zeros is replaced by a composition $\mathbf{r} = [r_1, r_2, \dots, r_D]$ without zeros as follows [13]:

$$r_j = \begin{cases} \delta, & \text{if~} x_j = 0, \\ \left(1 - \frac{z\delta}{\sum x_j}\right) x_j, & \text{if~} x_j > 0, \end{cases} \tag{1}$$

where z is the number of zeros counted in $\mathbf{x}$ and δ is the imputed value for part x_j computed as follows:

doi.org/10.32388/Z69TVG.2

$$\delta = \frac{\left(\min_{t,x} D_{t,x}\right)/2}{\sum_{x=0}^{110} D_{t,x}}, \forall D_{t,x} > 0. \tag{2}$$

Subsequently, $\mathbf{r}$ is multiplied by $\sum_{x=0}^{110} D_{t,x}$ to obtain a new set of death counts without zeros, which are then used to calculate mortality rates for ages below 80 [13]. Combining these with the mortality rates for older ages, a smoothed and imputed set of observed mortality rates is ready to be used for constructing life tables.

The country-specific life tables are constructed from the preprocessed mortality rates separately for males and females using the LifeTable function from the MortalityLaws package [36]. The average number of years lived by individuals dying within the age interval $[x, x+1)$, denoted as a_x, is assumed to be 0.5 for all single-year ages, except age 0 [13][33]. For age 0, a_0 for each country/region is computed as the average of a_0 values derived from the range of infant mortality rates m_0 during the training period (1983-2010), as outlined in Table 2.1 [37]. This approach is similar to that employed by the HMD [33], incorporating averaging to allow for country-specific and gender-specific adjustments, thereby reflecting infant mortality more accurately than using a fixed a_0 value across all populations.

The life table radix is assumed to be unity, ensuring that the $d_{t,x}$ fall within a standard simplex [38][9][13]. For visualisation purposes, $d_{t,x}$ values are multiplied by 100,000, a commonly used radix in demographic research [33][15][26].

Gender	m_0 range	Formula
Male	$[0, 0.02300)$	$\frac{1}{N}\sum_{t=1}^{N}(0.14929 - 1.99545 m_{t,0})$
	$[0.02300, 0.08307)$	$\frac{1}{N}\sum_{t=1}^{N}(0.02832 + 3.26021 m_{t,0})$
	$[0.08307, \infty)$	0.29915
Female	$[0, 0.01724)$	$\frac{1}{N}\sum_{t=1}^{N}(0.14903 - 2.05527 m_{t,0})$
	$[0.01724, 0.06891)$	$\frac{1}{N}\sum_{t=1}^{N}(0.04667 + 3.88089 m_{t,0})$
	$[0.06891, \infty)$	0.31411

Table 1. Formulas for computing a_0 based on m_0[37].

Note. $N = 28$ *is the length of the training set.*

2.2. The CLR transformation

A positive compositional data vector typically satisfies a unit sum constraint and lies within a sample space called the standard simplex [39][24], defined by

$$\mathbb{S}^D = \left\{ \mathbf{x} = [x_1, \ldots, x_D] \in \mathbb{R}^D \mid x_i > 0, \sum_{i=1}^{D} x_i = 1 \right\}. \tag{3}$$

Positive points can be mapped to the simplex $\mathbb{S}^D$ using the closure operator [39][25] defined as:

$$C : (0, \infty)^D \to \mathbb{S}^D, \tag{4a}$$

$$C[\mathbf{x}] = \left[\frac{x_1}{\sum_{i=1}^{D} x_i}, \ldots, \frac{x_D}{\sum_{i=1}^{D} x_i} \right]. \tag{4b}$$

When the simplex is equipped with Aitchison geometry and its associated operations, it is referred to as the Aitchison simplex which forms a vector space [21]. Some key features of Aitchison geometry [21][13] [39] are:

$$\text{Perturbation:} \quad \mathbf{x} \oplus \mathbf{y} = C[x_1 y_1, \ldots, x_D y_D], \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \tag{5a}$$

$$\text{Power transformation:} \quad \alpha \odot \mathbf{x} = C[x_1^{\alpha}, \ldots, x_D^{\alpha}], \forall \mathbf{x} \in \mathbb{S}^D, \alpha \in \mathbb{R}, \tag{5b}$$

$$\text{Negative perturbation:} \quad \mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus (-1 \odot \mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \tag{5c}$$

$$\text{Inner product:} \quad \langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(\ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} \right), \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D, \tag{5d}$$

$$\text{Norm:} \quad \|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_a}, \forall \mathbf{x} \in \mathbb{S}^D, \tag{5e}$$

$$\text{Distance:} \quad d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a, \forall \mathbf{x}, \mathbf{y} \in \mathbb{S}^D. \tag{5f}$$

Since compositional data carry only relative proportions, Aitchison [21] introduced log-ratio based transformations, including the widely used CLR transformation. The CLR transformation [21] that maps the simplex onto a hyperplane passing through the origin of $\mathbb{R}^D$ is defined as:

$$\text{clr: } \mathbb{S}^D \to \mathbb{H} \subset \mathbb{R}^D, \tag{6a}$$

$$\mathbf{w} = \text{clr}(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ldots, \ln \frac{x_D}{g(\mathbf{x})} \right], \tag{6b}$$

$$\text{clr}^{-1}(\mathbf{w}) = C\left[e^{\mathbf{w}} \right], \tag{6c}$$

where $g(\mathbf{x}) = \left(\prod_{i=1}^{D} x_i \right)^{\frac{1}{D}}$ is the geometric mean of the composition.

doi.org/10.32388/Z69TVG.2

The CLR transformation is a one-to-one mapping between $\mathbb{S}^D$ and $\mathbb{R}^D$ under a zero-sum constraint. It also preserves distances where $d_a\langle\mathbf{x},\mathbf{y}\rangle = d_e(\mathrm{clr}(\mathbf{x}),\mathrm{clr}(\mathbf{y}))$ with d_e representing the Euclidean distance [39].

2.3. The α-transformation

Proposed by Tsagris et al. [24], the α-transformation is a one-parameter Box-Cox type power transformation that maps compositional data $\mathbf{x}$ from D-dimensional Aitchison simplex $\mathbb{S}^D$ to $(D-1)$-dimensional unconstrained real space $\mathbb{R}^{D-1}$. Given a compositional vector $\mathbf{x} \in \mathbb{S}^D$, the transformation and its inverse are expressed as:

$$A_\alpha : \mathbb{S}^D \to \mathbb{R}^{D-1}, \tag{7a}$$

$$\mathbf{z}_\alpha = A_\alpha(\mathbf{x}) = \frac{1}{\alpha}\mathbf{H}_D\left(D\mathbf{u}_\alpha(\mathbf{x}) - 1_D\right), \tag{7b}$$

$$A_\alpha^{-1}(\mathbf{z}_\alpha) = C\left[\left(\alpha\mathbf{H}'_D\mathbf{z}_\alpha + 1_D\right)^{1/\alpha}\right], \tag{7c}$$

where $\mathbf{z}_\alpha \in \mathbb{R}^{D-1}$, $\mathbf{u}_\alpha(\mathbf{x}) = \left[\frac{x_1^a}{\sum_{j=1}^D x_j^\alpha}, \ldots, \frac{x_D^a}{\sum_{j=1}^D x_j^\alpha}\right]$ is the power transformed vector in $\mathbb{S}^D$, $\mathbf{H}_D$ is the $(D-1)\times D$ Helmert sub-matrix and 1_D is the D-dimensional vector of ones. $\mathbf{H}_D$ is a standard orthogonal matrix obtained by removing the first row from the orthonormal Helmert matrix, reducing the dimensionality of transformed vector to $D-1$ [40].

The parameter $\alpha \in [0,1]$ can be tuned using criteria tailored to the type of analysis, such as the pseudo-R^2, profile log-likelihood or Kullback-Leibler divergence [24][29]. As noted by Tsagris et al. [24], when α approaches 1, it simplifies to a linear transformation. When $\alpha = 0$, it is equivalent to the isometric log-ratio (ILR) transformation, which requires the data to be free of zeros:

$$\mathrm{ilr}: \mathbb{S}^D \to \mathbb{R}^{D-1}, \mathrm{ilr}(\mathbf{x}) = \mathbf{H}_D\mathrm{clr}(\mathbf{x}). \tag{8}$$

2.4. Modelling and forecasting

Following the pipelines proposed by Bergeron-Boucher et al. [13] and Oeppen [9], a matrix $\mathbf{D}_{T\times(X+1)}$ consists of $d_{t,x}$ is constructed for each country/region and gender, where the T rows representing the years and $X+1$ columns representing the ages. In this study, the data are split into training and test sets using the commonly chosen 80:20 ratio [41], resulting in a training period from 1983 to 2010 and a test period from 2011 to 2018. Hence, each matrix $\mathbf{D}$ in the training set consists of 28 rows (1983–2010) and 111 columns (ages 0 to 110+). Each row of the compositional data sums up to the life table radix.

The matrix $\mathbf{D}$ is then centred by subtracting the column-specific geometric means α_x, resulting in matrix $\mathbf{F}$. Transformations are applied to the matrix $\mathbf{F}$ to allow the compositional data to vary freely in unconstrained real space, forming matrix $\mathbf{H}$. Subsequently, SVD is applied to the matrix $\mathbf{H}$ to estimate κ_t and β_x through a rank-K approximation. Although a rank-1 approximation is commonly used, higher-rank approximations are adopted when the variance explained by the first component is insufficient [13]. Based on the proportion of explained variance, $K = 7$ for females and $K = 4$ for males are deemed appropriate, as they each account for over 80% of the total variance on average. This leads to a total number of K series of estimated κ_t for each dataset.

An autoregressive integrated moving average (ARIMA) model is then fitted to each κ_t series with a forecast horizon of 8. The approximately linear trend of κ_t makes it suitable for forecasting using an ARIMA model. Although the random walk with drift has been shown to provide a good fit [5], prior research [13] demonstrated that the ARIMA (0,1,1) with drift performs well for most Western European countries. Therefore, this study considers two forecasting models, namely (i) the default model, ARIMA (0,1,1) with drift and (ii) the automatic ARIMA model [42], which selects the optimal order using a stepwise algorithmic procedure. The resulting matrix is denoted as $\mathbf{H}^*$.

Then, the inverse transformation is applied to convert the data back to the simplex, forming matrix $\mathbf{F}^*$. Matrix $\mathbf{D}^*$ which contains the forecast life table death counts $\hat{d}_{t,x}$ is obtained by adding back α_x to $\mathbf{F}^*$. Eventually, a complete forecast life table providing a full mortality profile for a population can be constructed using $\hat{d}_{t,x}$.

2.5. Model evaluation

Forecast accuracy for each model, fitted to data transformed using both methods, is evaluated on the test set using root mean squared error (RMSE) and mean absolute error (MAE). For each country/region and gender, the model yielding the lowest out-of-sample forecast error is selected for comparison. A thorough comparative analysis is then conducted to assess the impact of the α-transformation on predictive performance relative to the benchmark CLR transformation.

3. Application to Real Data

This section presents the results and provides an in-depth discussion comparing model performance using data with different transformations.

3.1. α-parameter tuning

Country/Region	**Optimal α**	**Average validation RMSE (%)**
Austria	0.2354	0.0495
Belgium	0.3268	0.0553
Bulgaria	0.1000	0.0727
Belarus	0.1000	0.0973
Switzerland *	0.0000	0.0550
Czechia	0.3358	0.0591
East Germany	0.5364	0.0727
West Germany	0.1103	0.0532
Denmark	0.1409	0.0992
Spain *	0.0000	0.0443
Estonia	0.2295	0.1320
Finland	0.0766	0.1101
France	0.1000	0.0611
England & Wales *	0.0000	0.0505
Northern Ireland	0.1919	0.1444
Scotland	0.1000	0.0665
Greece	0.1000	0.0614
Hungary	0.2208	0.0691
Ireland *	0.0000	0.1235
Iceland	0.1000	0.2731
Italy	0.5226	0.0407
Lithuania	0.1000	0.0903
Luxembourg *	0.0000	0.1768

Country/Region	Optimal α	Average validation RMSE (%)
Latvia	0.0365	0.0895
Netherlands	0.1000	0.0703
Norway	0.1000	0.0737
Poland	0.1803	0.0673
Portugal *	0.0000	0.0707
Slovakia	0.1000	0.0731
Slovenia	0.1001	0.0927
Sweden	0.1000	0.0633

Table 2. Optimal α values for female mortality data.

Note. Asterisks (*) denote countries/regions with optimal α = 0.

The values of α are chosen via cross-validation on a data-driven basis [29]. In order to determine the optimal α values for transforming $d_{t,x}$ of each country/region, an expanding window approach [41] is adopted. As mentioned in Section 2.4, data from year 1983 to 2010 serve as the training set, while the remaining eight years (2011-2018) form the out-of-sample test set.

To tune the α-parameter, the training set undergoes an additional split into sub-training and validation sets. Starting with an initial sub-training set of 15 years, the training window expands by one year at a time, while maintaining a fixed validation period of four years. This results in a total of ten iterations. For example, in the first iteration, the sub-training set spans 1983-1997, while validation covers 1998-2001. In the final iteration, the sub-training set spans 1983-2006, with validation covering 2007-2010. By progressively increasing the sub-training set size, this approach enhances generalisability and mitigates the risk of overfitting to a limited subset of data.

The flexibility of the α-parameter allows the transformation to adapt to different mortality patterns, but excessive flexibility may lead to overfitting, particularly if α is overly sensitive to small variations in the sub-training data. To prevent this, the optim function in R [30] is used to select optimal α values within the range of $[0, 1]$ that minimise the average RMSE in the validation set. In addition, a penalisation mechanism is applied to exclude α values that yield negative detection limits, preventing implausible negative estimates of $\hat{d}_{t,x}$.

Using female data as an example, Table 3.1 tabulates the α values chosen for the best model that yields higher forecast accuracy. Interestingly, only six countries/regions have an optimal α value of 0, suggesting that the intermediate α values are more suitable for female mortality forecasting than both EDA and LRA.

3.2. Forecasts of mortality: A case study on Italian female mortality

The country-specific model accuracy shows that the α-transformation performs comparably to, or better than, the CLR transformation in 22 countries/regions for females and 23 for males. Using the Italian female data as an example, Figures 1a to 1b illustrate the forecast $\hat{d}_{t,x}$ based on both α and CLR transformations for the last holdout year (2018) and for selected 20-year-gapped ages over the years, respectively.

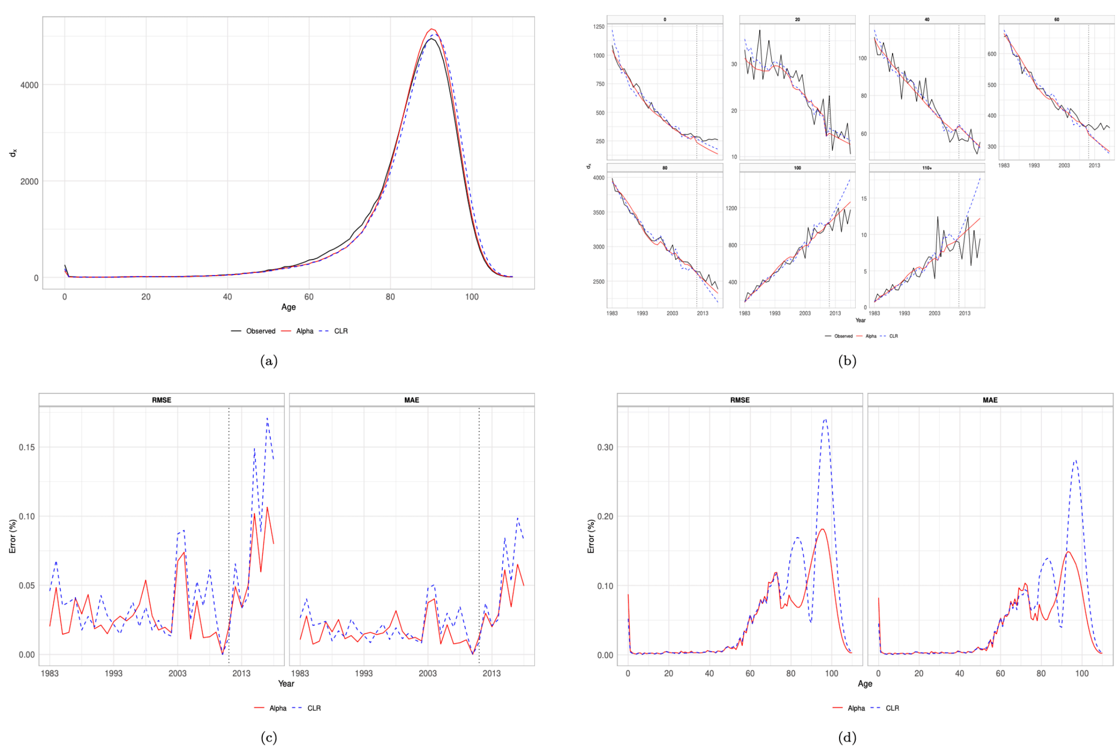

Figure 1. Forecast age-specific life table death counts (a) in 2018 and (b) for selected ages with forecast errors (c) over years and (d) by ages for females in Italy.

 doi.org/10.32388/Z69TVG.2

Based on Figure 1a, it is observed that the age-at-death distribution is negatively skewed and bimodal, with peaks in infancy that later shift towards older ages [15][43]. In general, infant mortality is primarily driven by genetic errors or infectious diseases while ageing becomes the main cause of mortality at older ages [43]. According to Abouzahr et al. [44], high mortality rates are typically found during infancy and reach their lowest levels between ages 5 and 14, before rising exponentially beyond age 35. There might also have some bumps during female reproductive ages, indicating premature mortality due to maternal deaths.

On the other hand, Figure 1b shows that death counts for infants and younger age groups up to age 80 generally exhibit a decreasing trend over the years, whereas ages 100 and above display an increasing trend, likely reflecting population ageing. Both transformations generate forecasts that broadly follow the historical trend. However, the α-transformation tends to produce smoother and more stable forecasts, particularly for age groups with volatile and low death counts such as ages 20 and 110+. In contrast, the CLR transformation appears more sensitive to variability in sparse data, likely due to the amplifying effect of log-ratio transformations on small values.

As the years progress in the test set, forecasts based on α-transformed data retain the trend patterns of $d_{t,x}$, similar to those predicted based on CLR-transformed data. Despite yielding similar forecasts, the α-transformation results in lower forecast errors as depicted in Figure 1c. Furthermore, forecasts from the α-transformed data are more accurate, especially beyond age 75, with a slight decrease in accuracy between ages 86 and 90 as shown in Figure 1d. More results for females and males can be found in the supplementary materials.

3.3. Comparison of mean forecast accuracy

3.3.1. Overall mean errors

Table 3.3.1 summarises the overall mean forecast errors for female mortality in both training and test sets. Both transformations yield comparable performance on the training and test sets, with the α-transformation producing slightly lower errors overall. Similar findings have been reported by Giacomello [25] and Shang and Haberman [26]. This is primarily due to the flexibility introduced by the α-parameter which allows it to better adapt to the underlying data structure [24], particularly the temporal changes in age-specific life table death counts [26].

Phase	RMSE (%)		MAE (%)	
	α	**CLR**	**α**	**CLR**
Train	**0.0619**	0.0621	**0.0348**	0.0349
Test	**0.0882**	0.0903	**0.0523**	0.0529

Table 3. Overall mean forecast errors for female mortality.

Note. Bold values indicate that the α-transformation yields comparable or lower errors than the CLR transformation.

3.3.2. Mean errors by country/region

Breaking down model forecast accuracy of female mortality at the country/region level, Table 3.3.2 presents RMSE and MAE values for both the training and test sets under the α and CLR transformations, along with their best ARIMA models. Notably, the ARIMA (0,1,1) with drift appears as the best-performing model in more than half of the countries/regions, suggesting its suitability for forecasting female mortality in Europe [13].

Country/Region	RMSE				MAE (%)				ARIMA model	
	Train		Test		Train		Test			
	α	CLR	α	CLR	α	CLR	α	CLR	α	CLR
Austria	0.0400	0.0400	**0.0962**	0.1251	0.0234	0.0232	**0.0546**	0.0688	Auto	Auto
Belgium	0.0418	0.0423	**0.0635**	0.0704	0.0243	0.0245	**0.0393**	0.0420	Default	Default
Bulgaria	0.0680	0.0681	**0.1200**	0.1192	0.0350	0.0350	**0.0673**	0.0670	Default	Default
Belarus	0.0521	0.0511	**0.2091**	0.2158	0.0315	0.0309	**0.1299**	0.1341	Default	Default
Switzerland *	0.0365	0.0365	**0.0485**	0.0485	0.0224	0.0224	**0.0299**	0.0299	Default	Default
Czechia	0.0503	0.0493	0.0648	0.0522	0.0269	0.0264	0.0362	0.0303	Auto	Auto
East Germany	0.0585	0.0545	**0.0704**	0.0800	0.0303	0.0286	**0.0416**	0.0443	Auto	Default
West Germany	0.0411	0.0448	**0.0584**	0.0600	0.0209	0.0227	**0.0344**	0.0350	Auto	Auto
Denmark	0.0562	0.0549	0.1383	0.1319	0.0333	0.0327	0.0775	0.0735	Auto	Default
Spain *	0.0310	0.0310	**0.0459**	0.0459	0.0167	0.0167	**0.0273**	0.0273	Default	Default
Estonia	0.0943	0.0938	0.0973	0.0881	0.0585	0.0583	0.0649	0.0579	Default	Default
Finland	0.0948	0.0927	**0.0966**	0.0995	0.0493	0.0482	**0.0572**	0.0587	Auto	Auto
France	0.0397	0.0404	**0.0669**	0.0819	0.0201	0.0203	**0.0368**	0.0435	Default	Default
England & Wales *	0.0323	0.0323	**0.0482**	0.0482	0.0172	0.0172	**0.0275**	0.0275	Default	Default
Northern Ireland	0.0897	0.0889	**0.0890**	0.0903	0.0516	0.0514	**0.0533**	0.0541	Auto	Auto
Scotland	0.0516	0.0515	**0.0561**	0.0560	0.0299	0.0299	**0.0362**	0.0362	Default	Default
Greece	0.0504	0.0503	**0.0680**	0.0679	0.0266	0.0267	**0.0404**	0.0403	Auto	Auto
Hungary	0.0501	0.0485	**0.0708**	0.0764	0.0273	0.0258	**0.0434**	0.0454	Auto	Default
Ireland *	0.0727	0.0727	**0.0727**	0.0727	0.0426	0.0426	**0.0432**	0.0432	Auto	Auto
Iceland	0.2006	0.2008	0.1905	0.1824	0.1112	0.1115	0.1012	0.0993	Auto	Auto
Italy	0.0325	0.0404	**0.0691**	0.1039	0.0161	0.0203	**0.0378**	0.0514	Default	Auto
Lithuania	0.0639	0.0637	0.0985	0.0965	0.0398	0.0398	0.0641	0.0631	Default	Default
Luxembourg *	0.1328	0.1328	**0.1862**	0.1862	0.0771	0.0771	**0.1077**	0.1077	Default	Default

doi.org/10.32388/Z69TVG.2

Country/Region	RMSE				MAE (%)				ARIMA model	
	Train		Test		Train		Test			
	α	CLR	α	CLR	α	CLR	α	CLR	α	CLR
Latvia	0.0749	0.0750	0.1418	0.1401	0.0466	0.0467	0.0869	0.0859	Default	Default
Netherlands	0.0389	0.0388	**0.0458**	0.0480	0.0215	0.0214	**0.0275**	0.0285	Default	Default
Norway	0.0570	0.0570	0.0511	0.0502	0.0321	0.0322	0.0307	0.0292	Default	Default
Poland	0.0404	0.0459	0.0837	0.0732	0.0204	0.0236	0.0501	0.0429	Default	Auto
Portugal *	0.0439	0.0439	**0.0539**	0.0539	0.0243	0.0243	**0.0338**	0.0338	Default	Default
Slovakia	0.0632	0.0638	0.1115	0.1044	0.0328	0.0332	0.0620	0.0581	Default	Default
Slovenia	0.0762	0.0754	**0.0807**	0.0904	0.0434	0.0429	**0.0521**	0.0574	Auto	Auto
Sweden	0.0449	0.0448	**0.0397**	0.0396	0.0259	0.0259	**0.0252**	0.0250	Auto	Auto

Table 4. Mean forecast errors of the best ARIMA models for female mortality in each country/region.

Note. The default model refers to ARIMA (0,1,1) with drift. Asterisks (*) denote countries/regions where the optimal α value is 0, while bold values indicate cases where the α-transformation yields comparable or lower mean forecast errors than the CLR transformation in the test set.

Generally, the α-transformation results in comparable or lower forecast errors than the CLR transformation in 22 countries/regions for females. It is also worth noting that when $\alpha = 0$, as observed in Switzerland, Spain, England & Wales, Ireland, Luxembourg and Portugal, the transformation reduces to the ILR transformation, resulting in the similar accuracy as the CLR transformation. This is in accordance with Shang and Haberman's [26] findings that both ILR and CLR transformations perform similarly in terms of point forecast accuracy for Australian gender-specific $d_{t,x}$ within functional CoDA framework.

3.3.3. Mean errors over years

Figure 2 visualises the trend of mean forecast errors for female mortality over years in the training and test sets. The α-transformation performs comparably to the CLR transformation, yielding slightly lower mean forecast errors, especially during the later years of the forecast horizon. In addition, it is worth highlighting that the extremely small errors up to 10^{-18} in the last fitting year, i.e. year 2010, are due to the jump-off adjustment. This occurs because the last observed $d_{t,x}$ are used as jump-off point, ensuring continuity between the observed and forecast life table death counts [10][45].

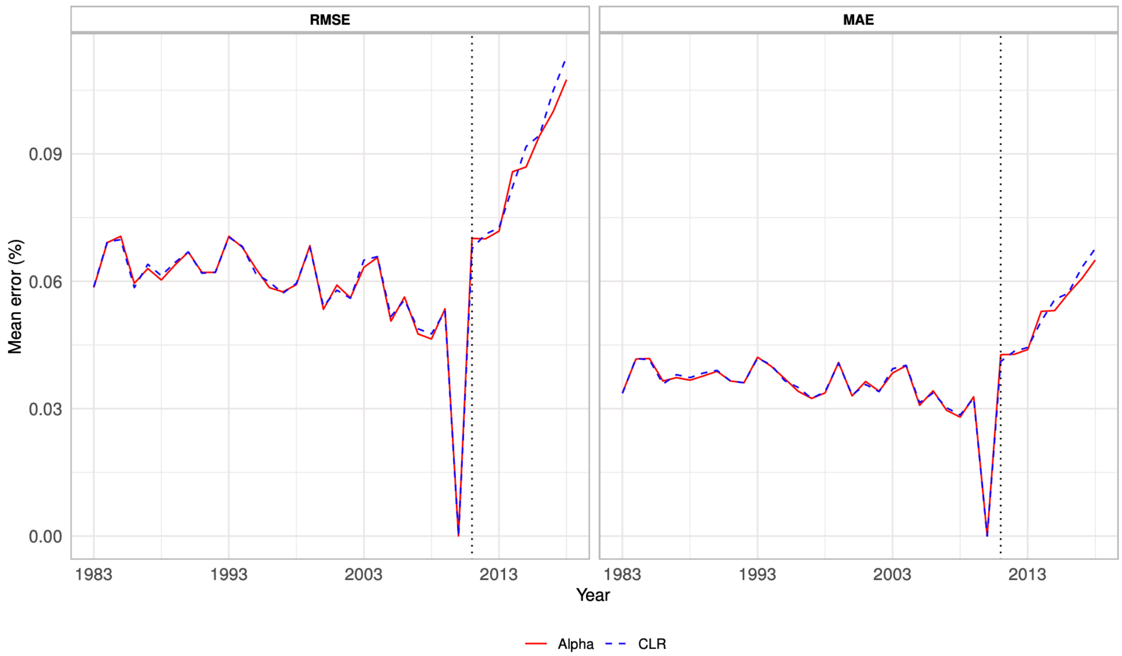

Figure 2. Mean forecast errors for female mortality over years across countries/regions.

When the widely used LC model was first introduced, Lee and Carter [5] noted that using fitted values as jump-off rates would not perfectly match the data in the jump-off year, leading to a discontinuity between observed and forecast log mortality rates. Although they claimed that such discontinuity affects only rates that are absolutely low with little impact on life expectancy, Lee and Carter [5] suggested that the jump-off error could be fixed by setting a_x equal to the most recently observed age-specific log mortality rates. However, this approach has a drawback that may deteriorate the goodness of fit over the remaining fitting period.

Lee and Miller [10] later highlighted the presence of significant jump-off bias, which requires adjustment to improve short-term forecast accuracy, especially when using low-rank approximations that tend to incur greater approximation error [46]. Bell [46] and Lee and Miller [10] discovered that a jump-off adjustment using observed jump-off rates can eliminate the bias and achieve a more accurate forecast in

 doi.org/10.32388/Z69TVG.2

forecasting period, as also proven by Stoeldraijer et al. [45]. Following this, the errors between $d_{t,x}$ and $\hat{d}_{t,x}$ reported here are nearly zero, as the fitted values are adjusted to match the observed values [10].

3.3.4. Mean errors by age

Figure 3 illustrates the mean forecast errors for female mortality by age across countries/regions measured using RMSE and MAE. Both the α-transformed and CLR-transformed data result in comparable forecast accuracy, although the former tends to exhibit slightly higher errors around ages 80 to 90. However, this situation does not persist as the α-transformation shows a clear advantage over the CLR transformation at older ages, particularly between ages 91 and 100. This result aligns with previous findings that the α-transformation can be an alternative to the CLR transformation for compositional mortality forecasting [25][26].

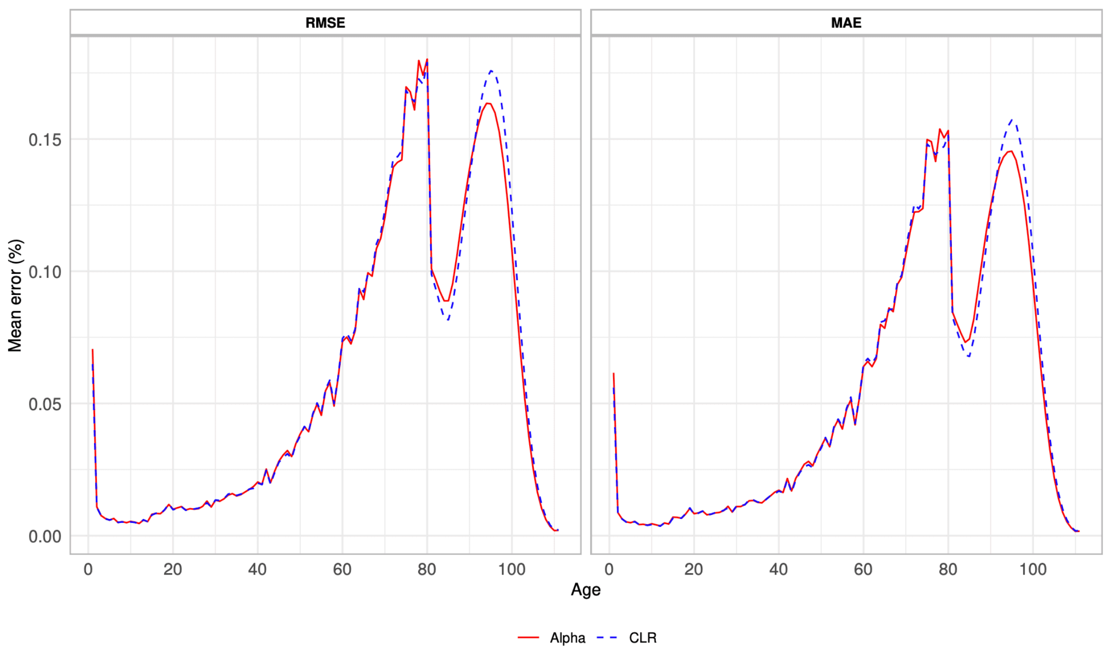

Figure 3. Mean forecast errors for female mortality by age across countries/regions.

4. Conclusion

This study investigates the use of the α-transformation as an alternative to the commonly used CLR transformation within the CoDA framework under a non-functional data setting for mortality forecasting. Unlike in the classical LC model, life table death counts are employed as the mortality measure. As these

doi.org/10.32388/Z69TVG.2

data are subject to a summability constraint, they are inherently compositional and should be transformed into an unconstrained real space prior to forecasting. Using age-specific life table death counts by gender across selected European countries/regions, a comparative analysis is conducted between the two transformations, evaluated from several perspectives, including overall mean forecast errors, errors by country/region, errors across years and errors by age.

The results show that models fitted to the α-transformed data perform comparably to those using the CLR transformation across most countries/regions, with slightly improved forecast accuracy observed in certain cases. This finding is consistent with earlier studies on the α-transformation within the functional CoDA framework [25][26]. The advantage of the α-transformation is particularly evident at older ages, where it yields noticeably improved accuracy for ages with low death counts. This improvement is largely attributed to its flexibility in bridging EDA and LRA [27], as the α-parameter can take intermediate values between 0 and 1, allowing for better data adaptation. In this study, optimal α values for most datasets fall within the intermediate range, determined by minimising the average validation RMSE during the parameter tuning phase.

A key limitation of this study lies in the dataset used. Since forecast accuracy is highly data-dependent, factors such as the selection of countries/regions, the fitting period and the level of age group aggregation can all influence model performance. In this context, high-quality data with sufficiently large volume is crucial to ensure consistent and reproducible results. Moreover, the study does not fully exploit the strengths of the α-transformation. A notable advantage of this transformation is its ability to handle compositional data containing zeros. However, due to the need for a fair comparison with the CLR transformation, zero values in the dataset were imputed, limiting the potential benefit of using the α-transformation.

Future directions of this study may include exploring alternative power transformation techniques that accommodate zeros, such as the chiPower transformation [47], to enhance predictive performance. Further work could also examine the performance of various transformations within a functional coherent CoDA framework that captures common trends across countries. For practical applications, particularly in the pension and insurance industries, it may be valuable to estimate annuity prices at different ages using forecasts of life table death counts, as demonstrated by Shang and Haberman [15].

Statements and Declarations

Conflicts of interest

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Universiti Malaya Research Excellence Grant (UMREG 2.0) 2024 [UMREG044-2024].

Data availability

All data used in this manuscript are available from the Human Mortality Database (www.mortality.org).

References

1. [a, b] *Cairns AJG, Blake D, Dowd K (2006). "A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration." J Risk Insur.* ***73****(4):687–718.*
2. [^] *Organisation for Economic Co-operation and Development (2024). "Pensions at a Glance Asia/Pacific 2024." OECD Publishing, Paris.*
3. [^] *Pollard JH (1988). "Projection of Age-Specific Mortality Rates." Popul Bull UN.* ***21/22****:55–69.*
4. [^] *Booth H, Ticker L (2008). "Mortality Modelling and Forecasting: A Review of Methods." Ann Actuar Sci.* ***3****(1-2):3–43.*
5. [a, b, c, d, e, f] *Lee R, Carter LR (1992). "Modeling and Forecasting U. S. Mortality." J Am Stat Assoc.* ***87****(419):659–671.*
6. [a, b, c] *Basellini U, Camarda CG, Booth H (2023). "Thirty Years On: A Review of the Lee-Carter Method for Forecasting Mortality." Int J Forecast.* ***39****(3):1033–1049.*
7. [a, b, c] *Hyndman RJ, Ullah MS (2007). "Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach." Comput Stat Data Anal.* ***51****(10):4942–4956.*
8. [a, b, c] *Li N, Lee R (2005). "Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method." Demography.* ***42****(3):575–594.*
9. [a, b, c, d, e, f, g] *Oeppen J (2008). "Coherent Forecasting of Multiple-Decrement Life Tables: A Test Using Japanese Cause of Death Data." In European Population Conference 2008, Barcelona, Spain, pp. 1–25.*

10. a, b, c, d, e *Lee R, Miller T (2001). "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality." Demography.* ***38****(4):537–549.*

11. ^ *Atance D, Claramunt MM, Varea X, Aburto JM (2024). "Convergence and Divergence in Mortality: A Global Study from 1990 to 2030." PLoS One.* ***19****(1):e0295842.*

12. ^ *King G, Soneji S (2011). "The Future of Death in America." Demogr Res.* ***25****:1–38.*

13. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o *Bergeron-Boucher M-P, Canudas-Romo V, Oeppen J, Vaupel JW (2017). "Coherent Forecasts of Mortality with Compositional Data Analysis." Demogr Res.* ***37****(17):527–566.*

14. a, b *Kjærgaard S, Ergemen YE, Kallestrup-Lamb M, Oeppen J, Lindahl-Jacobsen R (2019). "Forecasting Causes of Death by Using Compositional Data Analysis: The Case of Cancer Deaths." J R Stat Soc Ser C Appl Stat.* ***68****(5):1351–1370.*

15. a, b, c, d, e, f *Shang HL, Haberman S (2024). "Weighted Compositional Functional Data Analysis for Modeling and Forecasting Life-Table Death Counts." J Forecast.* ***43****(8):3051–3071.*

16. ^ *Jong PD, Marshall C (2007). "Mortality Projection Based on the Wang Transform." ASTIN Bull.* ***37****(1):149–161.*

17. ^ *Pascariu MD, Canudas-Romo V, Vaupel JW (2018). "The Double-Gap Life Expectancy Forecasting Model." Insur Math Econ.* ***78****:339–350.*

18. a, b, c, d, e *Bergeron-Boucher MP, Kjærgaard S, Oeppen J, Vaupel JW (2019). "The Impact of the Choice of Life Table Statistics When Forecasting Mortality." Demogr Res.* ***41****:1235–1268.*

19. ^ *Canudas-Romo V (2010). "Three Measures of Longevity: Time Trends and Record Values." Demography.* ***47****(2):299–312.*

20. ^ *Aburto JM, Villavicencio F, Basellini U, Kjærgaard S, Vaupel JW (2020). "Dynamics of Life Expectancy and Life Span Equality." Proc Natl Acad Sci USA.* ***117****(10):5250–5259.*

21. a, b, c, d, e, f, g, h *Aitchison J (1986). The Statistical Analysis of Compositional Data. Chapman & Hall, London, UK.*

22. ^ *Wilmoth JR (1995). "Are Mortality Projections Always More Pessimistic When Disaggregated by Cause of Death?" Math Popul Stud.* ***5****(4):293–319.*

23. ^ *Shang HL, Haberman S (2020). "Forecasting Age Distribution of Death Counts: An Application to Annuity Pricing." Ann Actuar Sci.* ***14****(1):150–169.*

24. a, b, c, d, e, f, g, h, i *Tsagris M, Preston S, Wood ATA (2011). "A Data-Based Power Transformation for Compositional Data." In 4th International Workshop on Compositional Data Analysis, J. J. Egozcue, R. Tolosana-Delgado, and M. I. Ortego (Eds.), Girona, Spain, pp. 1–8.*

25. a, b, c, d, e, f *Giacomello S (2024). "Functional α-Transformations for the Statistical Analysis of Density Functions." Master's thesis, Politecnico di Milano.*

26. a, b, c, d, e, f, g, h *Shang HL, Haberman S (2024). "Forecasting Age Distribution of Life-Table Death Counts Via α-Transformation." Scand Actuar J.* ***2025****(4):387–403.*

27. a, b, c, d *Tsagris M, Preston S, Wood ATA (2016). "Improved Classification for Compositional Data Using the α-Transformation." J Classif.* ***33****:243–261.*

28. a, b *Tsagris M, Alenazi A, Stewart C (2023). "Flexible Non-Parametric Regression Models for Compositional Response Data with Zeros." Stat Comput.* ***33****(5):106.*

29. a, b, c *Tsagris M (2015). "Regression Analysis with Compositional Data Containing Zero Values." Chil J Stat.* ***6****(2):47–57.*

30. a, b *R Core Team (2024). "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria.*

31. ^ *HMD Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France), 2024; dataset available at www.mortality.org.*

32. ^ *Hyndman RJ, Booth H, Tickle L, Maindonald J, Wood S, R Core Team (2023). "Demography: Forecasting Mortality, Fertility, Migration and Population Data." R package version 2.0.*

33. a, b, c, d, e, f *Wilmoth JR, Andreev K, Jdanov D, Glei DA, Riffe T, Boe C, Bubenheim M, Philipov D, Shkolnikov V, Vachon P, Winant C, Barbieri M (2021). "Methods Protocol for the Human Mortality Database." 6th ed.*

34. ^ *Thatcher AR, Kannisto V, Vaupel JW (1998). The Force of Mortality at Ages 80 to 120. Vol. 5. Odense University Press, Odense.*

35. a, b *Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V (2003). "Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation." Math Geol.* ***35****(3):253–278.*

36. ^ *Pascariu MD (2024). "MortalityLaws: Parametric Mortality Models, Life Tables and HMD." R package version 2.1.0.*

37. a, b *Andreev EM, Kingkade WW (2015). "Average Age at Death in Infancy and Infant Mortality Level: Reconsidering the Coale-Demeny Formulas at Current Levels of Low Mortality." Demogr Res.* ***33****(13):363–390.*

38. ^ *Aitchison J (1994). "Principles of Compositional Data Analysis." In Multivariate Analysis and Its Applications, T. W. Anderson, K. T. Fang, and I. Olkin (Eds.), IMS Lecture Notes Monograph Series, Vol. 24, pp. 73–81.*

39. a, b, c, d *Clarotto L, Allard D, Menafoglio A (2022). "A New Class of α-Transformations for the Spatial Analysis of Compositional Data." Spat Stat.* ***47****:100570.*

40. ^*Lancaster HO (1965). "The Helmert Matrices." Am Math Mon.* **72***(1):4–12.*

41. a, b*Hyndman RJ, Athanasopoulos G (2021). Forecasting: Principles and Practice. 3rd ed. OTexts, Melbourne, Australia.*

42. ^*Hyndman RJ, Khandakar Y (2008). "Automatic Time Series Forecasting: The Forecast Package for R." J Stat Softw.* **27***(3):1–22.*

43. a, b*Robertson HT, Allison DB (2012). "A Novel Generalized Normal Distribution for Human Longevity and Other Negatively Skewed Data." PLoS One.* **7***(5):e37025.*

44. ^*Abouzahr C, Mikkelsen L, Rampatige R, Lopez A (2012). "Mortality Statistics: A Tool to Enhance Understanding and Improve Quality." Pac Health Dialog.* **18***(1):247–270.*

45. a, b*Stoeldraijer L, Duin CV, Wissen LV, Janssen F (2018). "Comparing Strategies for Matching Mortality Forecasts to the Most Recently Observed Data: Exploring the Trade-Off Between Accuracy and Robustness." Genus.* **74***:16.*

46. a, b*Bell WR (1997). "Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates." J Off Stat.* **13***(3):279–303.*

47. ^*Greenacre M (2024). "The ChiPower Transformation: A Valid Alternative to Logratio Transformations in Compositional Data Analysis." Adv Data Anal Classif.* **18***:769–796.*

Declarations

Funding: This work was supported by the Universiti Malaya Research Excellence Grant (UMREG 2.0) 2024 [UMREG044-2024].

Potential competing interests: No potential competing interests to declare.